STAT C206B: Topics in Stochastic Processes

Feynman Liang*
Department of Statistics, UC Berkeley

Last updated: February 18, 2020

Contents

1 Lecture 8

2020-02-13

Current setup

- $\mathcal{P}$ probability measures on $(X, \mathcal{X})$
- $\tilde{X} = X \times X \times \cdots$, $\tilde{\mathcal{X}}$ product σ-field
- $\tilde{P} = \{\tilde{\pi} = \pi \otimes \pi \otimes \cdot : \pi \in \mathcal{P}\}$
- Cylinder cets $C(E_1^{i_1}, \ldots, E_p^{i_p}) = \prod_{i=1}^{\infty} F_i$ where $F_{i_r} = E_r$ and $F_i = X$ for $i \notin \{i_1, \ldots, i_p\}$
- $\tilde{S}$ exchangeable probability measures on $\tilde{X}$

Heading towards $\tilde{P}$ equal to extreme points of $\tilde{S}$.

Theorem 1

{thm:equality-implies-extreme}

Let $\sigma \in \tilde{S}$ such that

$$\sigma C(E_1, \ldots, E_n, E_1, \ldots, E_n) = [\sigma C(E_1, \ldots E_n)]^2$$

for all $n \in \mathbb{N}$ and sets $(E_1)^n \subset \mathcal{X}$. Then σ is an extreme point of $\tilde{S}$.

Proof. By contradiction, suppose $\sigma \in \tilde{S}$ is not extreme. Then $\sigma = \alpha\sigma' + (1-\alpha)\sigma''$ for $\alpha \in (0,1)$ and $\sigma' \neq \sigma'' \in \tilde{S}$. Since all probability measures on $\tilde{\mathcal{X}}$ are determined by values on cylinder sets (monotone class, Dynkin π-λ), there exists cylinder set $B = C(E_1, \ldots, E_n)$ such that $\sigma' B \neq \sigma'' B$. Let

$$A = C(E_1, \ldots, E_n, E_1, \ldots, E_n)$$

Then in view of **??**

$$\sigma A = \alpha\sigma' A + (1-\alpha)\sigma'' A \geq \alpha(\sigma' B)^2 + (1-\alpha)(\sigma'' B)^2$$

By Jensen's inequality

$$[\alpha\sigma' B + (1-\alpha)\sigma'' B]^2 < \alpha(\sigma' B)^2 + (1-\alpha)(\sigma'' B)^2$$

Hence

$$\sigma A > [\alpha\sigma' B + (1-\alpha)\sigma'' B]^2 = (\sigma B)^2$$

so that strict inequality holds, a contradiction. □

*feynman@berkeley.edu

Definition 2

For a probability measure π on a σ-algebra Υ of subsets of a set Y, let $E \in \Upsilon$ be such that $\pi E \neq 0$. Then the set function π_{vertE} defined on $F \in \Upsilon$ by the relation

$$\pi_{|E}(F) = \pi(E \cap F)/\pi E$$

is a probability measure on Υ, called the *conditional probability* given E.

Lemma 3

{lem:indep-cond-prob} *$E, F \in \Upsilon$ are independent iff $\pi_{|E}(F) \neq \pi_{|F}(E)$.*

Theorem 4

$\tilde{P}$ is the set of extreme points of $\tilde{S}$

Proof. If $\tilde{\pi} \in \tilde{P}$ is an IID product measure, then we have equality in Theorem 1 and hence an extreme point.

For the reverse inclusion, suppose $\sigma \in \tilde{S} \setminus \tilde{P}$. Since σ is exchangeable, it cannot be a product measure otherwise $\sigma = \otimes_i \pi_i \Rightarrow \sigma = \otimes_i \pi$ for some π contradicting $\sigma \notin \tilde{P}$.

Thus, there must be some sets E, $F_1, \ldots, F_n \in \mathcal{X}$ such that

$$\sigma C(E, F_1, \ldots, F_n) \neq \sigma C(E) \sigma C(F_1, \ldots, F_n) \tag{1}$$ {{eq:witness-non-equality}}

Introduce the shift transformation: for $A \in \tilde{X}$ let

$$UA = \left\{a \in \tilde{X} : (a_2, a_3, \ldots) \in A\right\}$$

i.e. $UA = X \times A$. Then $A \mapsto UA$ maps $\tilde{\mathcal{X}}$ into (not in general surjective) $\tilde{\mathcal{X}}$ and $\sigma UA = \sigma A$ for all $A \in \tilde{\mathcal{X}}$ by exchangeability.

The non-equality Eq. (1) can be rephrased: there exits $B = C(F_1, \ldots, F_n) \in \tilde{\mathcal{X}}$ and $E \in \mathcal{X}$ such that

$$\sigma[C(E) \cap UB] \neq \sigma C(E) \cdot \sigma B$$

Hence, it is impossible that $\sigma C(E)$ or $\sigma C(E^c)$ vanish, otherwise both sides are equal to zero or σB. Thus, we can define set functions σ', σ'' such that for $A \in \tilde{\mathcal{X}}$

$$\sigma' A = \sigma_{|C(E)}(UA), \quad \sigma'' A = \sigma_{|C(E^c)}(UA)$$

σ' and σ'' are distinct elements of $\tilde{S}$ by Lemma 3, and furthermore by the law of total probability

$$\sigma = [\sigma C(E)]\sigma' + [1 - \sigma C(E)]\sigma''$$

Thus, $\sigma \in \tilde{S}$ and $\sigma \notin \tilde{P}$ implies σ is not an extreme point of $\tilde{S}$. □

1.1 Hausdorff moment problem

For $n \in \mathbb{N}$, let $\Pr_{n,\theta}$ be a family of distributions with finite expectation θ and variance $\sigma_n^2(\theta)$. Denote expectation

$$\mathbb{E}_{n,\theta}(u) = \int_{\mathbb{R}} u(x) \Pr_{n,\theta}(dx)$$

Lemma 5

Suppose u bounded continuous, $\sigma_n^2(\theta) \to 0$ for each θ. Then $\mathbb{E}_{n,\theta}(u) \to u(\theta)$ and convergence is uniform in every finite interval on which $\sigma_n^2(\theta) \to 0$ uniformly.

In other words, $\Pr_{n,\theta} \xrightarrow{w} \delta_\theta$ (because L^2 convergence implies convergence in probability implies convergence in distribution) hence $\mathbb{E}_{n,\theta}(u) \to u(\theta)$.

Proof. Clearly

$$|\mathbb{E}_{n,\theta}(u) - u(\theta)| \le \int_{\mathbb{R}} |u(x) - u(\theta)| \Pr_{n,\theta}(dx)$$

There exists δ depending on θ, ε such that $|x-\theta| < \delta \Rightarrow$ the integrand is $< \varepsilon$.

Outside of this neighborhood, the integrand is less than some constant M so by Chebyshev's inequality the probability carried by the region $|x-\theta| > \delta$ is less than $\sigma_n^2(\theta)\delta^{-2}$. Thus, the right side will be $< 2\varepsilon$ as soon as n is sufficiently large so that $\sigma_n^2(\theta) < \varepsilon\delta^2/M$. This bound on n is independent of $|theta$ if $\sigma_n^2(\theta) \to 0$ uniformly over this interval. ☐

Example 6

{eg:binom-uniform-approx}

Let $\Pr_{n,\theta}$ binomial distribution concentrated on points k/n for $k \in \{0, \dots, n\}$, then $\sigma_n^2(\theta) = \theta(1-\theta)n^{-1} \to 0$ and therefore

$$B_{n,u}(\theta) = \sum_{k=0}^{n} u\left(\frac{k}{n}\right)\binom{n}{k}\theta^k(1-\theta)^{n-k} \to u(\theta)$$

uniformly in $\theta \in [0,1]$. $B_{n,u}$ is called the *Bernstein polynomial of degree n corresponding to u.*
This is the proof of Weierstrass approximation theorem; we have a polynomial in θ converging uniformly to a bounded continuous function u.

Aside: weak convergence in probability theory is really weak-$*$ convergence, when restricted to compact spaces.

Hausdorff Moment Problem: Given (c_i), when can we tell that $c_k = \int_0^1 x^k \mu(dx)$ for some probability measure μ on $[0,1]$?

Definition 7

The *differencing operator* Δ, when applied to a countable sequene (a_i), is defined by $\Delta a_i = a_{i+1} - a_i$.

It produces a new sequence $\{\Delta a_i\}_i$, and

$$\Delta^2 a_i = \Delta a_{i+1} - \Delta a_i = a_{i+2} - 2a_{i+1} + a_i$$

Fill in

we notice a reciprocity relation between $\{a_i\}$ and $\{c_i\}$. Multiply ?? by $\binom{v}{r}c_r$ and sum over $r \in \{0, \dots, v\}$.

Fill in

$$\sum_{r=0}^{v} c_r \binom{v}{r}\Delta^r a_i = \sum_{j=0}^{v} a_{i+j}\binom{v}{j}(-1)^{v-j}\Delta^{v-j}c_j$$

To approximate higher derivatives, if we let $a_k = u(x+kh)$ for a function u, point x, and a span $h > 0$, then we define

$$\Delta_h u(x) = [u(x+h) - u(x)]/h$$

be the difference ratio (approximaiton to derivative), and more generally

$$\Delta_h^r u(x) = h^{-r}\sum_{j=0}^{r}\binom{r}{j}(-1)^{r-j}u(x+jh)$$

In particular, $\Delta_h^0 u(x) = u(x)$.

Return to Example 6. The LHS is aa polynomial, called the *Bernstien polynomial of degree n* corresponding to the given function u. Denote it by $B_{n,u}$.

Fill in

Definition 8

A sequence $\{c_k\}$ such that $(-1)^r\Delta^r c_k \ge 0$ for $r \in \mathbb{N}^*$ is called *completely monotone.*

Let $\Pr$ be a distribution on $[0,1]$ and $\mathbb{E}(u)$ the integral of u wrt $\Pr$. The kth moment is defined by

$$c_k = \mathbb{E}(X^k) = \int_{[0,1]} x^k \Pr(dx)$$

Successive differences shows

$$(-1)^r \Delta^r c_k = \mathbb{E}(\boldsymbol{X}^k (1-\boldsymbol{X})^r k)$$

so the moment sequence $\{c_k\}$ is completely monotone. Take u an arbitrary continuous function on $[0,1]$. Integrate Example 6 for the Bernstein polynomial $B_{n,u}$ wrt Pr. We get

$$\mathbb{E}B_{n,u} = \sum_{j=0}^{n} u(jh)\binom{n}{j}(-1)^{n-j}\Delta^{n-j}c_j = \sum_{j=0}^{n} u(jh)p_j^{(n)}$$

where $h = n^{-1}$ and $p_j^{(n)} = \binom{n}{j}(-1)^{n-j}\Delta^{n-j}c_j$. Plugging in $u = 1$ shows

$$1 = \sum_{j=0}^{n} p_j^{(n)}$$

This means that each n the $p_j^{(n)}$ define a distribution putting weight $p_j^{(n)}$ on the point $jh = j/n$, denoted

$$\begin{tikzcd}
u \arrow[r, dash] \arrow[dr, "\substack{\text{integrate wrt } \Pr_n, \\ \text{which puts mass } p_j^{(n)} \text{ on } j/n}"'] & B_{n,u} \arrow[d, "\text{integrate wrt Pr}"] \\
 & \mathbb{E}_n(u) = \mathbb{E}(B_{n,u})
\end{tikzcd}$$

{fig:commute-over-bnu}

Figure 1: commute-over-Bnu

by $\Pr_n$. . Fill in

So far $\{c_k\}$ was the moment sequence for Pr. Now let $\{c_k\}$ be an arbitrary monotone sequence and define analogously

$$p_j^{(n)} = \binom{n}{j}(-1)^{n-j}\Delta^{n-j}c_j$$

Notice this turns a completely monotone sequence $\{c_k\}$ into a probability sequence. By definition, these are nonnegative. We will show they add to c_0. By the reciprocit yformula

$$\sum_{j=0}^{n} u(jh)p_j^{(n)} = \sum_{r=0}^{n} c_r \binom{n}{r} h^r \Delta_h^r u(0)$$

For the constant function $u = 1$, the RHS reduces to c_0 and this proves the assertion.

Thus, any completely monotone sequence $\{c_k\}$ subject to the norming condition $c_0 = 1$ defines a distribution $\{p_j^{(n)}\}$. Fill in

Let u be a degree N polynomial. Since $h = 1/n$, $\Delta_h^r u(0) \to u^{(r)}(0)$. Furthermore $n(n-1)\cdots(n-r+1)h^r \to 1$ and the series $\sum_{r=0}^{n} c_r \binom{n}{r} h^r \Delta_h^r u(0)$ only contains terms up to $r = N$ hence at most $N+1$ terms. As $n \to \infty$, we conclude

$$\mathbb{E}_n(u) \to \sum_{r=0}^{N} \frac{c_r}{r!} u^{(r)}(0)$$

for every degree N polynomial.

In particular, when $u(x) = x^r$ we get $\mathbb{E}_n(X^r) \to c_r$. In other words,

Theorem 9 (*Hausdorff*)

The moments $\{c_r\}$ of a probability distribution on $[0,1]$ form a completely monotone sequence with $c_0 = 1$. Conversely, any arbitrarily completely monotone sequence with $c_0 = 1$ corresponds to a unique distributino on $[0,1]$.

We know that for any polynomial u, $\mathbb{E}_n(u)$ convergese to a finite limit. From Weierstrass approximation theorem, it follows the same is true for any function u continuous in $[0,1]$. Denote the limit $\mathbb{E}_n(u)$ by $\mathbb{E}(u)$.

Given $\{c_k\}$, we need to show there exists distribution Pr such that the limit $\mathbb{E}(u)$ coincides with the expectation of u wrt Pr. But this is immediate from the Riesz representation theorem.

Any completely monotone sequence $\{c_k\}$ with $c_0 = 1$ given by $c_k = \int_0^1 x^k \Pr(dx)$ for suitable unique Pr. Completely monotone sequenecs with $c_0 = 1$ form a convex set in LCTVS $\mathbb{R}^\infty$ with product topology whose extreme points are of the form $c_k = x^k$ for some $x \in [0,1]$. Show that this is true.

2 Lecture 9

2020-02-18

Notation from Kallenberg "Foundations of Modern Probability."

Underlying probability space (Ω, A, P).

$E^{\mathcal{F}}[\cdot] = E[\cdot \mid \mathcal{F}]$ conditional probability given $\mathcal{F}$ a sub-σ-field of $\mathcal{A}$

Definition 10

The *conditional probability* of an event $A \in \mathcal{A}$ given σ-field $\mathcal{F}$ is

$$P^{\mathcal{F}} A = E^{\mathcal{F}} \mathbb{1}_A$$

Equivalently

$$P[A \mid \mathcal{F}] = E[\mathbb{1}_A \mid \mathcal{F}], \quad A \in \mathcal{A}$$

Thus $P^{\mathcal{F}} A$ is the almost surely unique random variable in $L^1(\mathcal{F})$ satisfying

$$E[P^{\mathcal{F}} A; B] = P(A \cap B) \quad \forall B \in \mathcal{F}$$

$E[X \mid \mathcal{F}]$ is the a.s. unique $\mathcal{F}$-measurable RV Y such that $E[X\mathbb{1}_B] = E[Y\mathbb{1}_B]$ for all $B \in \mathcal{F}$.

- $P^{\mathcal{F}} A = PA$ a.s. iff $A \perp \mathcal{F}$ (denoting independent)
- $P^{\mathcal{F}} A = \mathbb{1}_A$ a.s. if A agrees a.s. with a set in $\mathcal{F}$, i.e. $P(A \Delta B) = 0$ for $B \in \mathcal{F}$
- Positivity of $E^{\mathcal{F}}$ implies $0 \le P^{\mathcal{F}} A \le 1$ a.s.
- Monotone convergence property gives

$$P^{\mathcal{F}} \bigcup_n A_n = \sum_n P^{\mathcal{F}} A_n$$

  a.s., for $A_i \in \mathcal{A}$ disjoint

We would like to have an assignment $A \mapsto (P^{\mathcal{F}} A)(\omega)$ that for each fixed ω is a probability measure (c.f. regular conditional distributions).

Definition 11

A *kernel* between two measurable spaces $(T, \mathcal{T})$ and $(S, \mathcal{S})$ is a function $\mu : T \times S \to \overline{\mathbb{R}_+}$ such that $\mu(t, B)$ is $\mathcal{T}$-measurable in $t \in T$ for fixed $B \in \mathcal{S}$ and a measure in $B \in \mathcal{S}$ for fixed $t \in T$.
μ is a *probability kernel* if $\mu(t, S) = 1$ for all t
Kernels on the basic probability space Ω are called *random measures.*

For fixed $B \in \mathcal{S}$, $t \mapsto \mu(t, B)$ as a function $(T, \mathcal{T}) \to (\overline{\mathbb{R}_+}, \mathcal{B})$ is measurable. For fixed $t \in T$, the set function $B \mapsto \mu(t, B)$ is a measure on $(S, \mathcal{S})$.

A random measure on $(S, \mathcal{S})$ is a map $\nu : \Omega \times \mathcal{S} \to \overline{\mathbb{R}_+}$ such that for each $B \in \mathcal{S}$, $\omega \mapsto \nu(\omega, B)$ is $\mathcal{A}$-measurable (i.e. $\nu(\cdot, B)$ is a random variable) and for each $\omega \in \Omega$, $B \mapsto \nu(\omega, B)$ is a measure on $(S, \mathcal{S})$ (i.e. $\nu(\omega, \cdot)$ is a measure).

Fix a σ-field $\mathcal{F} \subset \mathcal{A}$ and random element ξ in some measurable space $(S, \mathcal{S})$.

Definition 12

A *regular conditional distribution* of ξ, given $\mathcal{F}$, we mean a version of the function $P[\xi \in \cdot \mid \mathcal{F}]$ on $\Omega \times \mathcal{S}$ which is a probability kernel from $(\Omega, \mathcal{F})$ to $(S, \mathcal{S})$, hence an $\mathcal{F}$-measurable random probability measure on S.

If $\eta \in (T, \mathcal{T})$, we can talk about the rcd of ξ given η as a random measure of the form

$$\mu(\eta, B) = P[\xi \in B \mid \eta] \text{ a.s.}, \quad Bb \in \mathcal{S}$$

where μ is a probability kernel $T \to S$.

In the extreme cases ξ is $\mathcal{F}$-measurable or independent of $\mathcal{F}$, $P[\xi \in B \mid \mathcal{F}]$ has the regular version $\mathbb{1}\{\xi \in B\}$ or $P\{\xi \in B\}$ respectively.

Definition 13

$(S, \mathcal{S})$ is a *Borel space* if there exists a Borel subset $B \subset \mathbb{R}$ such that if we equip B with the trace σ-field

$$\mathcal{B}(B) = \{B \cap C : C \in \mathcal{B}(\mathbb{R})\} = \{A \in \mathcal{B}(\mathbb{R}) : A \subset B\}$$

then there is a bijection $f : S \to B$ that is measurable with a measurable inverse.

Example 14

Any complete seperable metric space (with its Borel σ-field) is a Borel space. For example, $\mathbb{N}^\infty$.

Theorem 15

For any Borel space S and measurable space T, let $\xi \in S$ and $\eta \in T$ be random elements. Then there exists a unique probability kernel μ from T to S satisfying $P[\xi \in \cdot \mid \eta] = \mu(\eta, \cdot)$ a.s., and μ is unique a.e. $\mathcal{L}(\eta)$.

Proof. By assumption of Borelness, wlog assume $S \in \mathcal{B}(\mathbb{R})$. For every $r \in \mathbb{Q}$ we may choose some measurable $f_r = f(\cdot, r) : T \to [0, 1]$ such that

$$f(\eta, r) = P[\eta \le r \mid \eta] \text{ a.s.}$$

□

Finish with Lebesgue-Stieltjes measures

Theorem 16 (*Disintegration*)

Fix measurable spaces S and T, σ-field $\mathcal{F} \subset \mathcal{A}$, random $\xi \in S$ such that $P[\xi \in \cdot \mid \mathcal{F}]$ has a regular version ν. Further consider $\mathcal{F}$-measurable random $\eta \in T$ and measurable $f : S \times T \to \overline{\mathbb{R}}$ with $\mathbb{E}|f(\xi, \eta)| < \infty$. Then

$$\mathbb{E}[f(\xi, \eta) \mid \mathcal{F}] \overset{as}{=} \int \nu(ds) f(s, \eta)$$

Proof. In the case when $\mathcal{F} = \sigma(\eta)$ and $P[\xi \in \cdot \mid \eta] = \mu(\eta, \cdot)$ for some probability kernel μ from T to S, then this becomes

$$E[f(\xi, \eta) \mid \eta] \overset{as}{=} \int \mu(\eta, ds) f(s, \eta)$$

Integrating, we get the commonly used formula

$$Ef(\xi, \eta) = E \int \nu(ds) f(s, \eta) = E \int \mu(\eta, ds) f(s, \eta)$$

If $\xi \perp \eta$, we can take $\mu(\eta, \cdot) \equiv \mathcal{L}(\xi)$ deterministic and the above reduces to the relation to previous lemma.

To prove the theorem, let $B \in \mathcal{S}$ and $C \in \mathcal{T}$. Use averaging property of conditional expectations to get □ ?????

Definition 17 (***Conditional independence***)

$\mathcal{F}_i, \ldots, \mathcal{F}_n, \mathcal{G} \subset \mathcal{A}$ sub-σ-fields are *conditionally indepenent* given $\mathcal{G}$ if

$$P^{\mathcal{G}} \bigcap_{k \leq n} B_k \overset{as}{=} \prod_{k \leq n} P^{\mathcal{G}} B_k$$

wehre $B_k \in \mathcal{F}_k$.
For infinite collections $\{\mathcal{F}_t\}_{t \in T}$, require the same property for every finite subcollection $\{\mathcal{F}_{t_i}\}_{i=1}^n$ with distinct indices.
Use $\perp_{\mathcal{G}}$ to denote pairwise conditional independence given $\mathcal{G}$.

Proposition 18 (***Conditional independence, Doob***)

For $\mathcal{F}, \mathcal{G}, \mathcal{H}$ σ-fields, $\mathcal{F} \perp_{\mathcal{G}} \mathcal{H}$ iff

$$P[H \mid \mathcal{F}, \mathcal{G}] \overset{as}{=} P[H \mid \mathcal{G}]$$

for all $H \in \mathcal{H}$.

Proof. Assuming above and using the chain and pull-out properties of conditional expectations, we get for $F \in \mathcal{F}$, $H \in \mathcal{H}$

$$\begin{aligned} P^{\mathcal{G}}(F \cap H) &= E^{\mathcal{G}} P^{\mathcal{F} \vee \mathcal{G}}(F \cap H) = E^{\mathcal{G}}[P^{\mathcal{F} \vee \mathcal{G}} H; F] \\ &= E^{\mathcal{G}}[P^{\mathcal{G}} H; F] = (P^{\mathcal{G}} F)(P^{\mathcal{G}} H) \end{aligned}$$

where $\mathcal{F} \vee \mathcal{G}$ is the smallest σ-field containing both $\mathcal{F}$ and $\mathcal{G}$. This shows $\mathcal{F} \perp_{\mathcal{G}} \mathcal{H}$.

Conversely, assume $\mathcal{F} \perp_{\mathcal{G}} \mathcal{H}$ and using the chain and pull-out properties, we get for any $F \in \mathcal{F}$, $G \in \mathcal{G}$, and $H \in \mathcal{H}$

$$E[P^{\mathcal{G}} H; F \cap G] = E[(P^{\mathcal{G}} F)(P^{\mathcal{G}} H); G] = E[P^{\mathcal{G}}(F \cap H); G] = P(F \cap G \cap H)$$

By a monotone class argument, this extends to

$$E[P^{\mathcal{G}} H; A] = P(H \cap A), \qquad A \in \mathcal{F} \vee \mathcal{G}$$

and the result follows by the averaging characterizataion of $P^{\mathcal{F} \vee \mathcal{G}} H$. □

From this result, we can conclude some further useful properties. Let $\bar{\mathcal{G}}$ denote the completion of $\mathcal{G}$ wrt $\mathcal{A}$, generated by $\mathcal{G}$ and $\mathcal{N} = \{N \subset A\}$. ??

Corollary 19

For any σ-fields $\mathcal{F}$, $\mathcal{G}$, and $\mathcal{H}$, we have

- *$\mathcal{F} \perp_{\mathcal{G}} \mathcal{H}$ iff $F \perp_{\mathcal{G}} (\mathcal{G}, \mathcal{H})$*
- *$\mathcal{F} \perp_{\mathcal{G}} \mathcal{F}$ iff $\mathcal{F} \subset \overline{\mathcal{G}}$*

Proof. By proposition ??, both relations are equivalent to

$$P[F \mid \mathcal{G}, \mathcal{H}] \overset{as}{=} P[F \mid \mathcal{G}], \quad F \in \mathcal{F}$$

□

Proposition 20 (***Chain rule***)

For any $\mathcal{G}$, $\mathcal{H}$, and $\mathcal{F}_i$, $i \in [n]$, TFAE

- *$H \perp_{\mathcal{G}} (\mathcal{F}_i)_i$*
- *$\mathcal{H} \perp_{\mathcal{G}, \mathcal{F}_1, \ldots, \mathcal{F}_n} \mathcal{F}_{n+1}$ for all $n \geq 0$*

In particular, we have the commonly used equivalence

$$\mathcal{H} \perp_{\mathcal{G}} (\mathcal{F}, \mathcal{F}') \Leftrightarrow \mathcal{H} \perp_{\mathcal{G}} \mathcal{F}, \mathcal{H} \perp_{\mathcal{G},\mathcal{F}} \mathcal{F}'$$

Definition 21

An *extension* of $(\Omega, \mathcal{A}, P)$ is a product space $(\hat{\Omega}, \hat{\mathcal{A}}) = (\Omega \times S, \mathcal{A} \otimes \mathcal{S})$ equipped with a probability measure $\hat{P}$ satisfying $\hat{P}(\cdot \times S) = P$.

Any random element $\xi \in \Omega$ can be regarded as a function on $\hat{\Omega}$, so we can replace ξ with $\hat{\xi}(\omega, s) = \xi(\omega)$ which clearly has the same distribution.

For extensions of this type, we can retain our original notation and write P and ξ instead of $\hat{P}$ and $\hat{\xi}$.

Lemma 22 (*Extension*)

Fix probability kernel μ between measurable S and T. let $\xi \in S$ be random element. Then there exists random $\eta \in T$ defined on some extension of Ω such that $P[\eta \in \cdot \mid \xi] \overset{as}{=} \mu(\xi, \cdot)$ and also $\eta \perp_{\xi} \zeta$ for all ζ on Ω.

Proof. Let $(\hat{\Omega}, \hat{\mathcal{A}}) = (\Omega \times T, \mathcal{A} \otimes \mathcal{T})$. Define $\hat{P}$ by

$$\hat{P}A = \int_{\Omega} P(d\omega) \int_{T} \mathbb{1}_A(\omega, t)\mu(\xi(\omega), dt) = E \int \mathbb{1}_A(\cdot, t)\mu(\xi, dt), \qquad A \in \hat{\mathcal{A}}$$

Clearly $\hat{P}(\cdot \times T) = P$ and $\eta(\omega, t) \equiv t$ on $\hat{\Omega}$ satisfies $\hat{P}[\eta \in \cdot \mid \mathcal{A}]\mu(\xi, \cdot)$. In particular, $\eta \perp_{\xi} \mathcal{A}$ by Proposition ?? hence $\eta \perp_{\xi} \zeta$. $\square$

Lemma 23

μ probability kernel S to Borel space T. Then exists measurable $f : S \times [0,1] \to T$ such that if $\theta \sim U(0,1)$ then $f(s, \theta)$ has distribution $\mu(s, \cdot)$ for every $s \in S$.

Will use this to prove the transfer principle:

Theorem 24

$\xi \overset{d}{=} \tilde{\xi}$, η random elements in S and

$$resp.THenexists$$

$\in T$ with $(\tilde{\xi}, \tilde{\eta}) \overset{d}{=} (\xi, \eta)$.